

CORRECTED VERSION

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
29 August 2002 (29.08.2002)

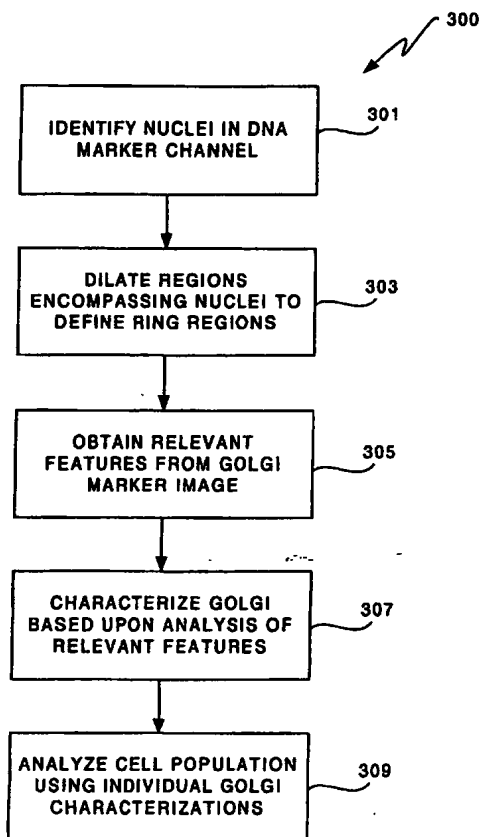
PCT

(10) International Publication Number  
WO 2002/067188 A3

- (51) International Patent Classification<sup>7</sup>: G01N 33/50, 15/14, G06K 9/00, 9/46
- (21) International Application Number: PCT/US2002/004492
- (22) International Filing Date: 14 February 2002 (14.02.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
09/792,012 20 February 2001 (20.02.2001) US
- (71) Applicant (for all designated States except US): CY-TOKINETICS, INC. [US/US]; 280 E. Grand Avenue, Suite 2, South San Francisco, CA 94080 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): CONG, Ge [CN/US]; 10944 San Pablo Avenue, Apartment 226, El Cerrito, CA 94530 (US). VAISBERG, Eugeni, A. [US/US]; 647 Pegasus Lane, Foster City, CA 94404 (US). WU, Hsien-Hsun [CN/US]; 1764 Wayne Circle, San Jose, CA 95131 (US).
- (74) Agent: WEAVER, Jeffrey, K.; Beyer Weaver & Thomas, LLP, P.O. Box 778, Berkeley, CA 94704 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI,

[Continued on next page]

(54) Title: IMAGE ANALYSIS OF THE GOLGI COMPLEX



(57) Abstract: Methods, code and apparatus analyze cell images to automatically identify and characterize the Golgi complex in individual cells. This is accomplished by first locating the cells in the image and defining boundaries of those cells that subsume some or all of the Golgi complex of those cells. The Golgi complex in the images typically have intensity values corresponding to the concentration of a Golgi component in the cell (e.g. a polysaccharide associated with the Golgi complex). The method/system then analyzes the Golgi components of the image (typically on a pixel-by-pixel basis) to mathematically characterize the Golgi complex of individual cells. This mathematical characterization represents phenotypic information about the cells' Golgi complex and can be used to classify cells. From this information, mechanism of action and other important biological information can be deduced.

WO 2002/067188 A3



SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU,  
ZA, ZW.

(88) Date of publication of the international search report:  
10 April 2003

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(48) Date of publication of this corrected version:  
1 April 2004

(15) Information about Correction:  
see PCT Gazette No. 14/2004 of 1 April 2004, Section II

**Published:**

— with international search report

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

## PATENT APPLICATION

## 5 IMAGE ANALYSIS OF THE GOLGI COMPLEX

BACKGROUND OF THE INVENTION

10 The present invention relates to methods and apparatus for characterizing a cell's condition based upon the state of its Golgi organelle. More specifically, the invention relates to image analysis methods and apparatus that rapidly characterize a cell based upon phenotypic characteristics of its Golgi complex.

15 The Golgi complex is a central and complex organelle within eukaryotic cells. It is involved in the sorting, modifying, and transporting of cellular molecules. It plays important roles in intracellular processes including endocytosis, exocytosis, and transport between cellular organelles. A detailed discussion of the role of the Golgi complex in cellular processes can be found in various treatises on cell biology. One example is Alberts et al. "Molecular Biology of the Cell" Garland Publishing, Inc., which is incorporated herein by reference for all purposes.

20 The Golgi complex can be an important indicator of the cellular effects caused by certain external agents. As a phenotypic characteristic, the Golgi complex has considerable value in drug discovery and fundamental biological research. For example, certain drugs and genetic modifications subtly affect transport pathways within a cell in ways that are manifest by the Golgi condition. Unfortunately, the value of the Golgi  
25 complex as an indicator has not been fully realized. This is because no simple consistent technique has been developed for characterizing the condition of Golgi complex in a high throughput manner.

SUMMARY OF THE INVENTION

30 This invention addresses the above need by providing methods, code and apparatus that analyze cell images to automatically identify and characterize Golgi in individual cells. The invention accomplishes this by first locating the cells in the image and defining subregions of those cells that subsume some or all of the Golgi of those cells. The Golgi

complex in the images typically have intensity values corresponding to the concentration of a Golgi component in the cell (e.g. a polysaccharide associated with the Golgi complex). The method/system then analyzes the Golgi components of the image to mathematically characterize the Golgi complex of individual cells. This mathematical  
5 characterization represents phenotypic information about the cells' Golgi complex and can be used to infer biological/physiological state of cells. From this information, mechanism of action and other important biological information can be deduced.

One aspect of this invention pertains to a method of analyzing an image of one or more cells. This method may be characterized by the following sequence of operations  
10 (typically implemented on a computing device): (a) identifying a region of the image subsuming some or all of the Golgi complex of a single cell; (b) within this region, automatically identifying the location of the Golgi complex; and (c) automatically mathematically characterizing the Golgi complex within the single cell. Preferably, the mathematical characterization is based upon (i) the Golgi complex location within the  
15 region and/or (ii) the concentration of a Golgi component within the region,

In addition to a basic characterization of the Golgi complex, which typically involves some basic morphological or statistical characterization, the process may automatically classify the Golgi complex in a category that at least partially distinguishes between normal Golgi and Golgi that is either diffuse or disperse or both diffuse and  
20 disperse. Such classification may be accomplished using a biological model, in the form of a neural network or regression model (e.g. a CART), for example.

Regarding the operation of identifying the region of the image subsuming some or all of the Golgi complex, the process may "segment" the image into multiple regions, each subsuming at least part of an individual cell. In a one example, segmentation involves  
25 identifying locations of nuclei in the cells, and then dilating the locations of the nuclei to subsume the locations of some or all of the Golgi complex.

Many different mathematical characterizations of the Golgi complex may be made. Preferably, all of them have biological relevance. Examples of types of mathematical characterizations include (i) an indicator of the peakedness of a histogram of at least one  
30 component of the Golgi complex, (ii) the texture of the Golgi complex and (iii) and the amount of Golgi complex in the region. As specific examples, the mathematical characterization of the Golgi complex include the kurtosis of intensity values obtained from the image, eigenvalues of a singular value decomposition of intensity values obtained from the image, and at least one of a mean and a standard deviation of intensity values  
35 obtained from the image.

Another aspect of this invention pertains to apparatus for automatically analyzing an image of one or more cells. Such apparatus may be characterized by the following features: (a) an interface configured to receive the image of one or more cells; (b) a memory for storing, at least temporarily, some or all of the image; and (c) one or more processors in communication with the memory and designed or configured to segment the image into discrete regions, each subsuming some or all of the Golgi complex in single cell. The processors may additionally characterize the Golgi complex of single cells by operating on the discrete regions. Still further, the processors may be designed or configured to classify the Golgi complex based upon such characterization of the Golgi complex. In one example, the classification distinguishes between normal Golgi and Golgi that is diffuse or disperse. The classification may make use of a biological model, in the form of a classification and regression tree, for example.

Still another aspect of the invention pertains to methods of producing a model for classifying cells based upon the condition of Golgi within the cells. Such method may be characterized by the following sequence: (a) receiving images of a plurality of cells of a training set; (b) analyzing the images to mathematically characterize the Golgi within the multiple cells from the training set; and (c) applying a modeling technique to the mathematical characterizations obtained in (b) to thereby produce the model. Typically, the training set will contain individual cells having Golgi in various states. The various states of Golgi in the cells of the training set may be produced, at least in part, by treatment with multiple exogenous agents such as drugs or drug candidates.

The process of analyzing the images mathematically may involve some of the operations outlined above such as segmentation, characterization, and classification. In one preferred approach, the modeling technique comprises generating a classification and regression tree.

Yet another aspect of the invention pertains to computer program products including machine-readable media on which are stored program instructions for implementing a portion of or an entire method as described above. Any of the methods of this invention may be represented, in whole or in part, as program instructions that can be provided on such computer readable media. In addition, the invention pertains to various combinations of data generated and/or used as described herein.

These and other features and advantages of the present invention will be described in more detail below with reference to the associated figures.

### BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-1D depict, in cartoon fashion, cells having Golgi in normal, dispersion, diffusion, and diffusion/dispersion states, respectively.

5 Figure 2 is a block diagram of an architecture and data flow for using multiple image analysis algorithms, each handling a different image of a microscopic field.

Figure 3 is a process flow diagram depicting one image analysis process for characterizing Golgi in accordance with this invention.

10 Figure 4A generally illustrates how an image of multiple cells may be segmented to provide separate representations of individual cells and thereby allow a cell-by-cell analysis.

Figure 4B generally depicts a representation of a cell as a binary object as identified using an edge finding routine.

Figures 5A and 5B depict a technique for "dilating" an image of a nucleus to create a perinuclear ring region where normal Golgi tend to localize.

15 Figure 6 is an intensity histogram showing peaked and flat curves corresponding to normal and diffuse Golgi, respectively.

Figure 7 is a cartoon depiction of cellular Golgi showing how normal and dispersed Golgi segregate inside and outside the perinuclear region, respectively.

20 Figure 8 is a sample classification and regression tree of the type that may be employed to classify Golgi in accordance with this invention.

Figure 9 is a block diagram of a computer system that may be used to implement various aspects of this invention such as the various image analysis algorithms of this invention.

25

### DETAILED DESCRIPTION OF INVENTION

A premise of this invention is that the shape and arrangement of Golgi in a cell provides valuable information about the cell's condition. If such condition results from treatment with a drug or other exogenous agent, then the Golgi complex may shed light on a mechanism of action.

In most normal cells, the Golgi complex is concentrated in a perinuclear region. In other words, normal Golgi is located in a cell's cytoplasm proximate to the nucleus. The chemical composition of the Golgi complex (and associated organelles and vacuoles) also affects the state of the Golgi within a cell's cytoplasm. Thus, when the particular drug or other exogenous agent affects a cell's cytoskeletal components and/or Golgi chemistry, this mechanism of action can be identified or suggested by examining the cell's Golgi complex.

In accordance with this invention, image analysis of the Golgi complex may be used to characterize the effects of any number of cellular stimuli. Stimuli of interest may include exposure to materials, radiation (including all manner of electromagnetic and particle radiation), forces (including mechanical, electrical, magnetic, and nuclear), fields, and the like. Examples of materials that may be used as stimuli include organic and inorganic chemical compounds, biological materials such as nucleic acids, carbohydrates, proteins and peptides, lipids, various infectious agents, mixtures of the foregoing, and the like. Other specific examples of agents include exposure to different temperatures, non-ambient pressure, acoustic energy, electromagnetic radiation of all frequencies, the lack of a particular material (e.g., the lack of oxygen as in ischemia), etc.

Turning now to Figures 1A, 1B, 1C, and 1D, the Golgi complex may exist in four separate and identifiable end states. Other detectable end states may also exist. For purposes of illustration, however, this document will focus on the four states depicted in Figures 1A – 1D. Figure 1A depicts a cell in an interphase stage of the cell cycle. Many cells have Golgi generally located as depicted, so long as they are not experiencing a major perturbation. As shown, a cell 101 includes a nucleus 103 and the Golgi apparatus 105. In this "normal" cell, the Golgi complex lies close to the nucleus 103 primarily on one side of the nucleus. A region of the cell where the Golgi complex resides is also proximate to the microtubule organizing center (MTOC), a structure within the cells from which the microtubules 107 radiate. Normal localization and morphology of the Golgi complex 105 is mediated by, among other factors, microtubules 107 and microtubule motors, proteins that drive the motion of various cellular components along microtubules.

Some changes in physiological conditions of the cell might cause fragmentation of the Golgi apparatus. Golgi in this condition is said to be "disperse." Figure 1B depicts, in cartoon fashion, a cell 109 in which Golgi fragments 111 are dispersed throughout the cytoplasm. This state of the Golgi complex may arise, for example, when the microtubule component of the cytoskeleton has been disrupted. Some drugs, for example, cause depolymerization of tubulin. When this occurs, the Golgi complex no longer remains

confined to tight bands in the MTOC. Rather, it disperses throughout the cytoplasm as depicted in Figure 1B. Other exogenous agents may disrupt microtubule motors. This too can cause dispersion of the Golgi complex. In general, exogenous agents that disrupt the microtubule cytoskeleton, either directly or indirectly, lead to dispersion of the Golgi complex.

Figure 1C depicts a third end state of the Golgi organelle. As shown there in cartoon fashion, a cell 113 has had its Golgi 115 spread throughout the cytoplasm without fragmentation. Here again the Golgi complex no longer remains confined to the perinuclear area. However, rather than forming discrete fragments it diffuses more or less evenly throughout the cells cytoplasm. Golgi in this condition is said to be in a "diffuse" state.

Golgi diffusion typically results when the dynamic barrier between the Golgi complex and endoplasmic reticulum compartments is disrupted to some degree. Actually, at this point, there may be no distinct Golgi. Rather the endoplasmic reticulum and Golgi have become one entity. This condition may result from a direct or indirect effect on the pathways between the endoplasmic reticulum and the Golgi complex. Any of a number of different exogenous and endogenous influences may cause this. In the end, the chemistry of the Golgi membrane changes significantly from its normal state.

The various states depicted in Figures 1A-1C represent "end states" in which the Golgi complex has reached an extreme. Obviously, depending upon the level of an effect causing cell disruption, the Golgi complex may exist in any one of these states only to a limited degree. Thus, for example, a cell may exist in a state between diffusion and dispersion as depicted in Figure 1D. As shown there, a cell 117 includes both diffuse Golgi components 115 and disperse Golgi components 111. In preferred embodiments, this invention can determine the degree to which a particular cell resides in any one of the 4 states depicted in Figures 1A- 1D.

Note also that the some of the influences may affect absolute amounts of Golgi components interacting with a particular Golgi marker. Some embodiments of this invention characterize the cell based upon the concentration of a marked Golgi component, which is affected by the degree to which an amount of a chemical constituent (or constituents) of the Golgi complex has changed from its normal state.

Note that the normal state depicted in Figure 1A is normal only for interphase cells, not for mitotic cells. Golgi naturally becomes dispersed and/or diffuse during mitosis. When a large number of dividing cells are present, these effects can obscure detection of



interesting changes to the Golgi complex that one might wish to detect. Therefore, the present invention may include operations that distinguish between mitotic and interphase cells. In the subsequent discussion, much of the characterization of the Golgi complex assumes that an algorithm has already limited consideration to interphase cells.

5

#### Preparation of the Image

Generally the images used as the starting point for the methods of this invention are obtained from cells that have been specially treated and/or imaged under conditions that contrast the cell's Golgi from other cellular components and the background of the image.

- 10 In the preferred embodiment, the cells are fixed and then treated with a labeling agent (i.e., a marker) that binds to one or more Golgi components and shows up in an image. Preferably, the chosen agent specifically binds to cell components that are enriched in the Golgi complex. The agent should provide a strong signal to show where it is concentrated in a given cell. To this end, the agent may be luminescent, radioactive, fluorescent, etc.
- 15 Various stains and fluorescent compounds may serve this purpose.

- The Golgi complex has numerous components that can be marked for generating images for use with this invention. For example, the Golgi contains specific proteins, lipids, and polysaccharides that can be marked for analysis. Some of these components occur in much higher concentration within the Golgi complex than within other cellular
- 20 components. In one preferred embodiment, the Golgi complex is identified using fluorescently labeled antibodies that bind specifically to a Golgi component. In a particularly preferred embodiment, the Golgi complex is identified by treatment with labeled Lens culinaris lectin (LC lectin). Lectins generally bind very specifically to certain polysaccharides. LC lectin binds to polysaccharide chains on proteins (proteoglycans)
- 25 having a N-acetyl-d-glucoseamine residue at the end of the polysaccharide chain. Alternatively one could use antibodies to proteins enriched in the Golgi complex, such as gp130, [beta]COP.

- Various techniques for preparing and imaging appropriately treated cells are described in U.S. Patent Applications 09/310,879, 09/311,996, and 09/311,890, previously
- 30 incorporated by reference. In the case of cells marked with a fluorescent material, a collection of such cells is illuminated with light at an excitation frequency. A detector is tuned to collect light at an emission frequency. The collected light is used to generate the image and highlights regions of high Golgi component concentration.

Sometime corrections must be made to the measured intensity. This is because the absolute magnitude of intensity can vary from image to image due to non-linearities in the image acquisition procedure and/or apparatus. Specific optical aberrations can be introduced by various image collection components such as lenses, filters, beam splitters, polarizers, etc. Other non-linearities may be introduced by an excitation light source, a broad band light source for optical microscopy, a detector's detection characteristics, etc. For example, some optical elements do not provide a "flat field." As a result, pixels near the center of the image have their intensities exaggerated in comparison to pixels at the edges of the image. A correction algorithm may be applied to compensate for this effect. Such algorithms can be easily developed for particular optical systems and parameter sets employed using those imaging systems. One simply needs to know the response of the systems under a given set of acquisition parameters.

#### General Algorithms for Image Analysis

Various algorithms can be used to analyze cell images that highlight Golgi. Certain specific examples will be described herein. In general, much of the relevant information associated with Golgi can be obtained by specifically analyzing interphase cells and ignoring, or treating separately, mitotic cells. Preferably, algorithms of this invention identify the normal, diffuse, and disperse states of Golgi within interphase cells. To this end, the algorithm may obtain certain morphological and/or statistical information about the Golgi complex from the image and use that information to classify the Golgi.

Most images of relevance will depict the Golgi complex as variations in intensity over position in the image, with higher intensity regions in the image corresponding to regions of the cell where the Golgi component exists in relatively high concentrations. Examples of morphological and statistical image characteristics that can be used to effectively characterize the Golgi include the "peakedness" of a histogram of pixel intensities, the texture of the Golgi regions of the image, the overall or total intensity obtained in the image, and the moments of the Golgi complex about a nucleus identified in a cell image.

Often it will be desirable to use an image analysis algorithm for Golgi in conjunction with image analysis algorithms for other components of a cell. In fact, in a preferred algorithm described herein, it is necessary to first identify the nucleus of a cell prior to identifying the Golgi complex.

Figure 2 depicts, from a global perspective, an algorithm (or group of algorithms) that characterizes a cell in terms of its nuclei, Golgi, and tubulin. As shown in Figure 2, a

global image analysis 200 begins with one or more images 201 as input. In this example, image 201 possesses three separate "channels." Each of these channels represents a different label that can be separately captured by the imaging apparatus and represented distinctly in an image of the cell or a field of cells. In a typical example, the separate  
5 labels are separate fluorophores having distinct emission frequencies. In a very specific example, cells are treated with a marker for DNA emitting in blue, a marker for the Golgi complex emitting in green, and a marker for tubulin emitting in red. As depicted in Figure 2, the image analysis algorithm 200 includes separate process branches: a nuclei analysis branch 203, a Golgi analysis branch 205, a tubulin analysis branch 207.

10 Each of these branches has its own separate algorithm for analyzing the image. As shown, the nuclei are analyzed with an algorithm 209, the Golgi complex are analyzed with an algorithm 211, and the tubulin is analyzed with a algorithm 213. Each of these algorithms act on information taken from the channel specific to the cell component of interest. The output of each of these algorithms is the combination of extracted features  
15 and, possibly (not necessarily), a separate classification or characterization of the associated cell component according to its state in a particular cell. As depicted, the output of algorithm 209 is the combination of identified individual nuclei (this is the part of the output that is later used by the Golgi algorithm), their morphological and intensity features, and a classification of the cell based upon its DNA or nucleus features. In a  
20 preferred embodiment, block 215 classifies the cell based upon its stage in the cell cycle. The output of algorithm 211 are Golgi features which are input to 217 for classification. Preferably, this classification is based upon (1) the type of Golgi arrangement (e.g. the degree of diffusion, dispersion, and normalcy) and/or (2) the concentration of one or more Golgi components (depicted as intensity of signal emitted by one or more Golgi markers).  
25 Finally, the output of tubulin algorithm 213 is the tubulin features which are input to 219 for classification. In one embodiment, this classification characterizes the overall shape of a cell.

Figure 3 depicts a process flow that may be employed to characterize Golgi appearing in images of cells. The process depicted in Figure 3 (identified by the reference  
30 300) may serve as a combination of algorithms 209 and 211 depicted in Figure 2, for example.

Initially, the individual cells, or at least the regions of those cells harboring the Golgi complex, are identified. Generally, this process is referred to as "segmentation." This can be accomplished in numerous ways. In the embodiment depicted in Figure 3, it is  
35 accomplished in operations 301 and 303. In this example, the image analysis tool initially

identifies the nucleus of each cell captured in the image under consideration. Since images from different channels can be well registrated, nucleus can be first identified in the DNA channel and then overlaid to the image under consideration. Identifying nucleus in the DNA channels may be accomplished using any of a number of nucleus  
5 identification/segmentation algorithms. As shown in Figure 2, the output of nucleus algorithm 209 is provided to Golgi algorithm 211.

After the nuclei have been identified at 301, the Golgi image analysis routine next defines a "ring region" around each nucleus. See 303. Generally, this step serves to define the perinuclear region. It will be described in detail with reference to Figures 4A and 4B.  
10 For now, it sufficient to understand that the purpose of operation 303 is to define a region that will encompass or subsume some or all of the Golgi complex in a normal interphase cell.

Within the regions defined in operation 303, the features of the Golgi complex may be identified using intensity versus position data. For example, each pixel in the ring  
15 regionis associated with a particular intensity, at the channel in which the Golgi complex is imaged. From such intensity data, various morphological and/or statistical features may be extracted in order to characterize the Golgi complex. See 305. These features will be described in more detail below. In one particularly interesting embodiment, the features of interest include kurtosis, standard deviation, mean, and singular value decomposition, all  
20 obtained from the intensities of the pixels in the ring region.

The statistical and morphological features can provide very useful information for characterizing Golgi complex. In themselves, however, these parameters have no biological meaning. In a preferred embodiment, algorithm 300 converts information contained in these parameters to biologically relevant classifications. This operation is  
25 depicted in block 307. In a preferred embodiment, operation 307 classifies all cells into one of four primary Golgi states: normal, diffuse, disperse, and diffuse/disperse.

Further biologically relevant information can be obtained by considering an entire population of cells, and their associated Golgi states. Thus, algorithm 300 concludes with an operation 309, in which a population analysis is conducted. In one example, the  
30 relative percentages of cells in any of the stages of cell cycle having Golgi in each of the four above-mentioned end states is computed from data on Golgi status and cell cycle stage of each individual cell. Note that statistics of the Golgi status may be computed independently for interphase and mitotic cells.

### Segmentation

One approach to segmentation is depicted in Figure 4A. As shown there, an image 401 includes a plurality of cell nuclei images 403 identified, for example, with a DNA-binding marker. Segmentation effectively converts image 401 into discrete images/representations for the DNA of each cell as shown at 405. In a preferred embodiment, this collection of representation 405 is provided as a mask providing intensity as a function of position for each cell nucleus in image 401.

Individual cell representations 405 may be extracted from image 401 by various image analysis procedures. Preferred approaches include edge finding routines and threshold routines. Some edge finding algorithms identify pixels at locations where intensity is varying rapidly. For threshold routines, pixels contained within the edges will have a higher intensity than pixels outside the edges. Threshold algorithms convert all pixels below a particular intensity value to zero intensity in an image subregion (or the entire image, depending upon the specific algorithm). The threshold value is chosen to discriminate between nuclei images and background. All pixels with intensity values above threshold in a given neighborhood are deemed to belong to a particular cell nucleus.

The concepts underlying thresholding are well known. A threshold value is chosen to extract those features of the image having intensity values deemed to correspond to actual cells (nuclei). Typically an image will contain various peaks, each having collections of pixels with intensity values above the threshold. Each of the peaks is deemed to be a separate "cell" or "nucleus" for extraction during segmentation.

An appropriate threshold may be calculated by various techniques. In a specific embodiment, the threshold value is chosen as the mode (highest value) of a contrast histogram. In this technique, a contrast is computed for every pixel in the image. The contrast may be computed as the intensity difference between a pixel and its neighbors. Next, for each intensity value (0-255 in an eight byte image), the average contrast over all pixels in the image is computed. The contrast histogram provides average contrast as a function of intensity. The threshold is chosen as the intensity value having the largest contrast. See "The Image Processing Handbook," Third Edition, John C. Russ 1999 CRC Press LLC IEEE Press, and "A Survey of Thresholding Techniques," P.K. Sahoo, S. Soltani and A.K.C. Wong, Computer Vision, Graphics, and Image Processing 41, 233-260 (1988), both of which are incorporated herein by reference for all purposes.

In another specific embodiment, edge detection may involve convolving images with the Laplacian of a Gaussian filter. The operation performed on the image is given by

comprises a two-dimensional collection of locations or pixels. Each of these is represented by a small rectangular box. The pixels where the nucleus resides are shown with the reference letter "N." To define the ring region, the algorithm locates the nearest neighbor pixels about the edge of the nucleus. In one embodiment, this is accomplished by identifying the four nearest neighbor pixels for each pixel identified as belonging to the nucleus. Those neighboring pixels which themselves form part of the nucleus are not considered. In Figure 5B, these nearest neighbor pixels are identified by the reference letter "X." To provide the appropriate width for the ring region, this process may be repeated any number of times. In the example shown in Figure 5B, it is repeated once-to provide a total of two iterations. In the second iteration, each of the pixels considered in the first iteration is used to identify four nearest neighbors. In Figure 5B, the second iteration nearest neighbors are identified by the reference letter "O."

The exact number of iterations used to define the ring region can vary depending upon the number of parameters. First, one must decide whether the ring region should subsume only Golgi residing in the close perinuclear region or should subsume Golgi residing within a wider portion of the cell, or even the entire cell itself. Second, the desired width of the ring region may vary depending upon the magnification of the image, the resolution of the image, and other related properties of the image.

Note that the mechanism depicted in Figure 5B represents but a single approach to defining the ring region. Some other suitable techniques will be readily apparent to those of skill in the art. Further, other techniques may be employed to define the region within which the Golgi may be subsumed. For example, some techniques will not consider the DNA or nucleus of a cell. Rather, they will simply focus on the image of the pertinent Golgi component and then consider appropriate morphological and/or statistical parameters to distinguish the Golgi of one cell from the Golgi of another cell. Still further, other techniques will consider other non-Golgi features of an image to distinguish one cell from another. For example, some algorithms may distinguish cells based upon images of the plasma membrane, cytoskeleton, etc. U.S. Patent Application No. 09/792,013 (Atty. Docket No. CYTOP013), previously incorporated by reference, describes a technique that uses tubulin (or other cytoskeletal component) and DNA markers together to identify discrete cells. Regardless of which technique is employed, in the end, regions subsuming some or all of the Golgi within discrete cells must be separately identified. Then, the image of the appropriate Golgi marker can be analyzed to characterize the Golgi in accordance with this invention. Note that while separate analysis of Golgi regions in each individual cell is often preferable, this is not a necessary requirement of the algorithm. Upon creation of "ring" areas, all pixels in all of the rings can be analyzed simultaneously.

Here  $I$  is the intensity of a pixel,  $x$  and  $y$  are the coordinates in the original image,

$$f(x, y) = \iint I(x - u, y - v) g(u, v) du dv$$

and  $g$  is the Laplacian of a Gaussian filter  $\frac{x^2 + y^2 - 2\sigma^2}{2\pi\sigma^8} e^{-\frac{x^2 + y^2}{2\sigma^2}}$ . The zero-crossings of the resulting image  $f(x, y)$  are detected as edge points. The edge points are linked to form closed contours, thereby segmenting the relevant image objects. See The Image Processing Handbook, referenced above. Figure 4B depicts the transformation in a very course fashion. An original image 422 is convolved with the Laplacian of a Gaussian filter to give a new image 424 which contains positive and negative values at the component pixels. Note that this figure shows the mask of an object, not detected edges/contours. Wherever the sign changes in moving from one pixel to the next a contour results.

#### 10 Generation Of Ring Region

As indicated in the discussion of 3, it is often desirable to generate a ring about the cell nucleus (see block 303) in order to define a region where normal Golgi are likely to reside. This operation is depicted in cartoon fashion in Figure 5A. As shown there, an image of a cell's nucleus 503 is bounded by a ring region 505. Within ring region 505, a normal Golgi organelle 507 appears in the image. As explained above, normal Golgi typically reside in the perinuclear region, which is coextensive with ring region 505.

Note that the size of the ring region can be chosen to allow some or all of the normal Golgi to be subsumed. In a preferred embodiment, the size of ring region 505 is chosen to subsume all of the normal Golgi in a majority of interphase cells 507. Diffuse and dispersed Golgi typically are not confined to a perinuclear region. Therefore, ring region 505 may not, depending upon the ring width setting, subsume all the Golgi in these states. The algorithm can use this fact to distinguish, or help distinguish, normal Golgi from diffuse and dispersed Golgi. Assuming that the total "quantity of Golgi" is approximately consistent across normal, diffuse, and disperse states, then the "amount of Golgi" located within ring region 505 will be greatest in the case of the normal Golgi state. Because some significant fraction of the Golgi lies outside of ring region 505 in the case of diffuse and/or disperse Golgi states, the total amount of Golgi detected within region 505 will be less in these states. In preferred embodiments, this is not the primary method used to deduce the Golgi state. Other features of pixel distributions (described below) are used more typically.

The ring shaped region about the nucleus may be defined using various techniques. One suitable technique is depicted in Figure 5B. As shown there, an image segment 511

### Relevant Golgi Features

In order to effectively characterize the Golgi complex of individual cells, the present invention may make use of various parameters derived from Golgi complex images. These parameters may represent the shape, size, concentration, texture, and amount of Golgi components in an individual cell. In one preferred algorithm of the present invention, one set of parameters is used to characterize the "type" of Golgi and another parameter or set of parameters is used to characterize the amount of Golgi complex in an individual cell. The types of Golgi may include normal, diffuse, disperse, and disperse/diffuse as discussed above. Of course, other suitable processes may consider different and/or additional types of Golgi. The amount of Golgi may be determined from the local concentrations of a marked Golgi component summed over the entire region of interest (e.g. a ring region).

Note that while the discussion below is organized sequentially, the described features of different nature may be used simultaneously to classify Golgi states. The parameters chosen for use with this invention should allow discrimination between biologically relevant classifications. Further, they may provide a continuous measure of the degree to which an individual cell exhibits features of any of the classifications. As mentioned, particularly interesting parameters will relate to one or more of the following general features of the Golgi complex image: the shape (such as peakedness) of an intensity (marker concentration) histogram, the texture of the Golgi areas, the overall intensity (some of local marker concentrations) of the Golgi image, and the moment (including moments of different orders) of the Golgi complex about the nucleus. Another set of features for the Golgi complex are statistics computed from a histogram of the intensity of pixel intensities by angle from the center of the nuclei. The x-axis of this histogram is the angle of the vector connecting the center of the nuclei to the ring pixel. The y-axis is the sum of the pixel intensities at a given angle. When the histogram is unimodal the Golgi complex is likely classified as normal, if the histogram has many modes the Golgi complex is likely classified as dispersed and if the histogram is uniform the Golgi complex is likely classified as diffuse.

Regarding peakedness, Figure 6 depicts sample histograms for normal and diffuse Golgi. As shown in Figure 6, a histogram plots the pixel count (number of pixels in an image meeting a criteria) as a function of pixel intensity. Remember that the intensity is a measure of the local concentration of a marked component in the Golgi complex. A Golgi complex with high local concentrations of the marked Golgi component will have a



narrow intensity distribution as shown by curve 601. A diffuse Golgi complex will have a wider distribution as illustrated by curve 603.

Generally, normal Golgi, and to some degree dispersed Golgi, possess local regions of high concentration of marked Golgi component. Therefore, these types of Golgi will have histograms with relatively narrow distributions. In contrast, diffuse Golgi complexes have their marked components relatively evenly distributed throughout most or all of the ring area. Therefore, the histogram of such diffuse Golgi will have a relatively wide distribution. In Figure 6, curve 601 might represent the histogram of normal or dispersed Golgi, while curve 603 might represent a histogram of diffuse Golgi.

Another parameter of interest in characterizing Golgi is the texture. The texture may be characterized by a number of parameters such as the singular value decomposition of a matrix of intensity values representing the marked Golgi component. Generally the texture of an object relates to the size and shape of the granules or other components of an image. Various texture related features can help to discriminate between normal and dispersed Golgi and/or between dispersed and diffuse/dispersed states.

The overall intensity associated with an image of the Golgi complex can represent the total amount of particular Golgi component(s) within a given cell or region of a cell. Figure 7 illustrates how this concept can be used to distinguish between normal Golgi and dispersed Golgi using an algorithm that employs a ring region as described above. As shown in Figure 7, a cell 701 possesses normal Golgi while a cell 703 possesses dispersed Golgi. Each cell includes a nucleus 705 and a plasma membrane 707. In the case of cell 701, the Golgi complex (represented by reference number 709) resides entirely within a ring region 711. In this cell, all of its Golgi resides within ring region 711. Therefore, the total intensity of a Golgi marker found within ring region 711 will represent all of the Golgi in cell 701. In contrast, cell 703 has a comparable ring region 711' which subsumes only a fraction of the Golgi in that cell (709'). The remainder of the dispersed Golgi lies outside the ring region 711'. As a consequence, it may be expected that the total intensity of Golgi calculated for cell 703 (using ring region 711') will be less than the total intensity calculated for cell 701.

The biologically relevant features of the Golgi complex may be obtained from various mathematical parameters. Among the parameters of interest are the following: the total area of the ring region in which the Golgi marker is considered, the mean intensity of all pixels in the ring region, the standard deviation of the pixel intensities in ring region, the kurtosis of the pixel intensities in the ring region, and the second, third, and fourth largest eigenvalues obtained by a singular value decomposition of a matrix of pixels in the

ring. Also, sometimes, the Golgi "type" (normal, diffuse, dispersed, or diffuse and dispersed) is treated as a parameter. This is technically a characterization derived from parameters such as the SVD eigenvalues, the kurtosis, etc. In a particularly preferred embodiment, the following four relevant parameters are used to characterize the Golgi complex: mean, standard deviation, kurtosis, and singular value decomposition.

Singular value decomposition and kurtosis can be derived from digitized images using conventional approaches. Kurtosis is given by  $m_4/(\text{std})^4$ , where std is the standard deviation of intensity of the pixels in the ring region in the image and  $m_4$  is given by the following expression:

$$m_4 = \frac{1}{N} \sum_{i=1}^N (x_i - \tilde{x})^4$$

In this expression, N is the number of pixels in the ring region in the image,  $x_i$  is the pixel intensity, and  $\tilde{x}$  is the mean pixel intensity. The kurtosis is particularly useful in distinguishing diffuse Golgi from normal or dispersed Golgi. Particularly, higher values of kurtosis suggest a higher degree of the Golgi "diffusion."

The singular value decomposition is technique that operates on a matrix of pixel intensity values arranged in their relative spatial positions as in the image. Preferably, these pixel intensities are obtained from a ring region as described above. The relevant values are obtained by multiplying the matrix with its transpose and obtaining the eigenvalues. The singular value decomposition provides information about the texture of the Golgi complex in the image. Thus, it is useful in distinguishing different states of the Golgi complex.

One aspect of the present invention is the ability to estimate the amount of Golgi (or a component of the Golgi) in a given cell or region of a cell. This allows one to characterize cells based upon relative or absolute quantity of a particular Golgi component within some region of the cell such as the perinuclear region. In general, an agent applied to highlight Golgi should emit a signal that is proportional to the amount of agent that has bound to a Golgi component. Thus, the amount of signal (usually indicated by the signal intensity summed or integrated over the entire region of interest) provides a direct indication of the quantity of a Golgi component present. When this is the case, the present invention allows one to obtain an accurate measure of the amount of Golgi in a given cell or cell region.

Techniques for estimating the quantity of DNA in a cell nucleus are described US Patent Application No. 09/729,754, previously incorporated by reference. These techniques may involve subtracting background signal and other image processing procedures. In general, such techniques apply to estimating the quantity of Golgi in a cell or region of a cell.

### Classification of Golgi

Optionally, the algorithms of this invention may specifically classify the Golgi complex into one or more biologically relevant classification based upon parameters such as those described above. As mentioned previously, one example of biologically relevant classifications include the four classifications depicted in Figures 1A- 1D. In the context of the general process flow depicted in Figure 3, the biologically relevant classification takes place at block 307. This operation takes as its inputs morphological, textual, and/or statistical parameters and provides as an output the biological relevant classifications. It accomplishes this using an appropriate model provided in the form of a neural network, linear or non-linear mathematical expression, a tree or graph, and the like. As previously, mentioned, one preferred approach employs a classification and regression tree.

Classification and regression trees (CART) are well known tools for classifying objects. They are described in Brieman et al., (1984) Classification and Regression Trees. Monterey: Wadsworth and Brooks/Cole. Figure 8 depicts one example of a classification and regression tree (CART) 800. In this example, there are three input parameters and four possible classifications. The input parameters are identified as,  $f_1$ ,  $f_2$ , and  $f_3$ . The classifications are denoted  $C_1$ ,  $C_2$ ,  $C_3$ , and  $C_4$ . In this hypothetical example, the tree initially considers the input parameter  $f_3$ . If the value of  $f_3$  is less than 0.1, the tree branches in one direction. If, on the other hand, the value of  $f_3$  is greater than or equal to 0.1, the tree branches in the opposite direction. Assuming that the value of  $f_3$  is less than 0.1, the tree next requires that the parameter  $f_2$  be considered. In this example, if the value of  $f_2$  is less than 100, the model classifies the Golgi complex in  $C_1$ . On the other hand, if the value of  $f_2$  is greater than or equal to 100 the model classifies the Golgi complex in  $C_2$ . Similarly, if the value of  $f_3$  is found to be greater than or equal to 0.1, then the value of the input parameter  $f_1$  is considered. Should the value of  $f_1$  be less than 3, the tree classifies the Golgi complex in  $C_3$ . Finally, if the value of  $f_1$  is found to be greater than or equal to 3, then the model classifies the Golgi complex in  $C_4$ .

Obviously, the parameters  $f_1$ ,  $f_2$ , and  $f_3$  will be relevant to classification of Golgi, in accordance with this invention. For example,  $f_1$  might correspond to the mean pixel intensity of the Golgi channel,  $f_2$  might correspond to the standard deviation of the pixel intensity values of the Golgi channel, and  $f_3$  might correspond to the second, third, and/or fourth singular value decomposition of a matrix of pixel intensity values in the ring region. The classification  $C_1$  through  $C_4$  might correspond to the Golgi complex states depicted in Figures 1A through 1D.

Classification models may be generated using numerous techniques. These include regression analyses (e.g., techniques for generating CARTs), neural networks, maximum likelihood/mixture models, etc. In general, any model requires a valid training set containing numerous samples that span the range of likely cases that will be encountered in practice. The samples should span a wide range of classification types, including types that vary in degree between extreme end cases (e.g., the diffuse/disperse Golgi state). In addition, the samples should include a wide range of input parameters. Each sample will have an ascribed classification and clearly defined input parameters. In this case, it will typically be necessary for an experienced scientist to classify images based upon their Golgi state. These classifications together with the relevant parameters (e.g. mean, standard deviation, kurtosis, and singular value decomposition) are provided to a tool for generating the appropriate classification model.

To maximize the predictive accuracy of the model, it may be appropriate to generate separate models for specific ranges of conditions. Typically, a separate model will be generated for each different cell line to be considered, because the appearance of the Golgi complex varies widely from cell line to cell line. Further, settings associated with the image analysis apparatus may cause significant variations in the image analysis. Therefore, separate models may be appropriate for different image analysis settings such as magnification, illumination intensity and the like. As noted, the Golgi complex may be imaged using various components contained within the Golgi complex. For example, some components may be detected with labeled lectins and other markers may be detected with labeled antibodies. Each of these separate component-marker combinations may deserve its own model, assuming more than one marker is used for the Golgi complex.

As indicated above, certain embodiments of the invention do not necessarily employ a classification operation. In these embodiments, the algorithm simply employs the parameters generated a block 305 in order to draw conclusions about a population of cells. Or alternatively the analysis tool may generate coefficients for biologically relevant features such as diffusion. In one example, the kurtosis of the Golgi image (or some

derivation from kurtosis) serves as a coefficient for diffusion. Further, the system could output both the biologically relevant classification of end states and the particular parameters used to generate these classifications.

#### Analysis of Populations of Cells

5       As mentioned, the most useful biological information often comes from images of a population of cells. The population may comprise a very limited range of variations, such as a single cell treated with a single drug at a single concentration. Or it may comprise a diverse set of variations such as multiple cell lines, each treated with multiple concentrations of a drug (or even multiple drugs believed to operate via a single  
10       mechanism of action).

Note that many images used with the present invention will contain many discrete cells. The images often show several hundreds of cells, although the actual number depends on the application. Such cells may collectively comprise the population of interest. Or multiple wells or a subset of a single well may comprise the population.

15       Regardless of how the population is defined, it should preferably contain a relevant sample of cells exposed to a condition of interest. From this population, conclusions about a given stimulus' effect on the Golgi complex can be drawn. The Golgi complex in the population can be characterized by the percent of cells in each primary category of Golgi end state or by a distribution of Golgi specific parameters across the  
20       population or by some other measure of Golgi parameters and/or classes across the population. As mentioned, interphase cells typically hold the most relevant information about an effect of a stimulus on the Golgi complex. This is because unperturbed interphase cells typically have normal Golgi of the type depicted in Figure 1A. Deviations from this normal state, such as illustrated in Figures 1B, 1C, and 1D, provide clues about  
25       the effect of a stimulus. Mitotic cells, in contrast, do not exhibit the "normal" Golgi state even in the absence of an additional stimulus and instead exhibit the diffuse and/or disperse state. Therefore, it is useful to characterize cells based on their position in the cell cycle – at least as either mitotic or interphase.

30       As explained in U.S. Patent Application No. 09/729,754, previously incorporated by reference, an image of a cells nucleus (DNA) can be analyzed to distinguish between the G<sub>1</sub>, S, and G<sub>2</sub>, and M phases of a cell. Briefly, the total amount of DNA in a cell nucleus can indicate whether an interphase cell is in the G<sub>1</sub>, S, or G<sub>2</sub> phase. Another parameter or grouping of parameters can discriminate between mitotic and interphase cells. One example of such parameter is the variance in intensity exhibited by an indicator

of DNA concentration. One example of a group of parameters is average pixel intensity and area of the nucleus. In this second example, the average pixel intensity is the average of the pixel-based intensities of the nucleus and "area" is the total area of the nucleus.

5 In a particularly preferred embodiment, the cells of a population are first characterized using marked DNA. This classification at least distinguishes between mitotic and interphase cells. Then, the same cells are characterized using marked Golgi. Particular attention is paid to the Golgi complex of the interphase cells.

#### Software/Hardware

10 Generally, embodiments of the present invention employ various processes involving data stored in or transferred through one or more computer systems. Embodiments of the present invention also relate to an apparatus for performing these operations. This apparatus may be specially constructed for the required purposes, or it may be a general-purpose computer selectively activated or reconfigured by a computer program and/or data structure stored in the computer. The processes presented herein are  
15 not inherently related to any particular computer or other apparatus. In particular, various general-purpose machines may be used with programs written in accordance with the teachings herein, or it may be more convenient to construct a more specialized apparatus to perform the required method steps. A particular structure for a variety of these machines will appear from the description given below.

20 In addition, embodiments of the present invention relate to computer readable media or computer program products that include program instructions and/or data (including data structures) for performing various computer-implemented operations. Examples of computer-readable media include, but are not limited to, magnetic media such as hard disks, floppy disks, and magnetic tape; optical media such as CD-ROM disks;  
25 magneto-optical media; semiconductor memory devices, and hardware devices that are specially configured to store and perform program instructions, such as read-only memory devices (ROM) and random access memory (RAM). The data and program instructions of this invention may also be embodied on a carrier wave or other transport medium. Examples of program instructions include both machine code, such as produced by a  
30 compiler, and files containing higher level code that may be executed by the computer using an interpreter.

Figure 9 illustrates a typical computer system that, when appropriately configured or designed, can serve as an image analysis apparatus of this invention. The computer system 900 includes any number of processors 902 (also referred to as central processing

units, or CPUs) that are coupled to storage devices including primary storage 906 (typically a random access memory, or RAM), primary storage 904 (typically a read only memory, or ROM). CPU 902 may be of various types including microcontrollers and microprocessors such as programmable devices (e.g., CPLDs and FPGAs) and unprogrammable devices such as gate array ASICs or general purpose microprocessors. As is well known in the art, primary storage 904 acts to transfer data and instructions unidirectionally to the CPU and primary storage 906 is used typically to transfer data and instructions in a bi-directional manner. Both of these primary storage devices may include any suitable computer-readable media such as those described above. A mass storage device 908 is also coupled bi-directionally to CPU 902 and provides additional data storage capacity and may include any of the computer-readable media described above. Mass storage device 908 may be used to store programs, data and the like and is typically a secondary storage medium such as a hard disk. It will be appreciated that the information retained within the mass storage device 908, may, in appropriate cases, be incorporated in standard fashion as part of primary storage 906 as virtual memory. A specific mass storage device such as a CD-ROM 914 may also pass data uni-directionally to the CPU.

CPU 902 is also coupled to an interface 910 that connects to one or more input/output devices such as video monitors, track balls, mice, keyboards, microphones, touch-sensitive displays, transducer card readers, magnetic or paper tape readers, tablets, styluses, voice or handwriting recognizers, or other well-known input devices such as, of course, other computers. Finally, CPU 902 optionally may be coupled to an external device such as a database or a computer or telecommunications network using an external connection as shown generally at 912. With such a connection, it is contemplated that the CPU might receive information from the network, or might output information to the network in the course of performing the method steps described herein.

In one embodiment, the computer system 900 is directly coupled to an image acquisition system such as an optical imaging system that captures images of cells. Digital images from the image generating system are provided via interface 912 for image analysis by system 900. Alternatively, the images processed by system 900 are provided from an image storage source such as a database or other repository of cell images. Again, the images are provided via interface 912. Once in the image analysis apparatus 900, a memory device such as primary storage 906 or mass storage 908 buffers or stores, at least temporarily, digital images of the cell. Typically, the cell images will show locations where Golgi, and possibly DNA, exists within the cells. In these images, local values of a Golgi image parameter (e.g., radiation intensity) correspond to amounts of a Golgi component at the locations within the cell shown on the image. With this data, the image

analysis apparatus 900 can perform various image analysis operations such as distinguishing between normal and diffuse and/or dispersed Golgi and estimating the amount of Golgi in a cell. To this end, the processor may perform various operations on the stored digital image. For example, it may analyze the image in manner that extracts values of one or more Golgi state parameters that correspond to a Golgi primary state and classifies the cell as either normal of diffuse, disperse or diffuse/disperse based upon the extracted values of the parameters. Alternatively, or in addition, it may estimate a total value of the Golgi image parameter taken over at least a perinuclear region of the cell.

### Examples

The following examples each involved generation and analysis of images obtained from SKOV3 cells treated with a Lens culinaris lectin marker. The cells were incubated with each drug for 24 hours before fixation and staining. In each case, ring regions were obtained by the Golgi segmentation algorithm described above. In addition, the Golgi complex within those ring regions was predicted, using a CART model as described above, as either normal, diffuse, dispersed, or diffuse and dispersed.

In a first example, SKOV3 cells that were treated with a low concentration (0.6 nanomolar) of the anti-microtubule drug Taxol. In this example, the Golgi complex remains normal as determined by the model. An expert independently examined the images and concluded that the Golgi complex was normal. In a second example, SKOV3 cells treated with a high concentration (2.6 micromolar) of microtubule-de polymerizing drug Nocodazole. Using the same model, the Golgi complex was found to be dispersed. And the expert independently confirmed that the Golgi complex was in fact dispersed. Finally, in a third example, SKOV3 cells were treated with a drug that disrupts transport into and out of the Golgi complex - Brefeldin A. The drug was provided in a concentration of 79 micromolar. In this case, the image analysis algorithm classified the Golgi complex as diffuse. The expert independently confirmed this.

### Conclusion

Although the above has generally described the present invention according to specific processes and apparatus, the present invention has a much broader range of applicability. In particular, the present invention is not limited to a particular kind of cell component, not just the Golgi complex. Thus, in some embodiments, the techniques of the present invention could provide information about many different types or groups of cellular organelles and components of all kinds. Of course, one of ordinary skill in the art would recognize other variations, modifications, and alternatives.



## CLAIMS

*What is claimed is:*

- 5     1.     A method, implemented on a computing device, of analyzing an image of one or more cells, the method comprising:
- (a) identifying a region of the image subsuming some or all of the Golgi complex of a single cell;
- (b) within said region, automatically identifying the location of the Golgi complex;
- 10    and
- (c) based upon at least one of (i) the Golgi complex location within the region and (ii) the Golgi concentration within the region, automatically mathematically characterizing the Golgi complex within the single cell.
- 15    2.     The method of claim 1 further comprising:
- (d) based upon the mathematical characterization, automatically classifying the Golgi complex in a category that at least partially distinguishes between normal Golgi and Golgi that is either diffuse or disperse or both diffuse and disperse.
- 20    3.     The method of claim 1, wherein the image comprises multiple cells, and wherein identifying the region of the image subsuming some or all of the Golgi complex comprises segmenting the image into regions, each subsuming at least part of an individual cell.
4.     The method of claim 3, wherein segmenting comprises identifying locations of
- 25    nuclei in the cells.
5.     The method of claim 4, wherein the nuclei are identified by identifying regions of the image where DNA is shown to concentrate.
- 30    6.     The method of claim 4, further comprising within the image dilating the locations of the nuclei to subsume the locations of some or all of the Golgi complex in the individual cells.
7.     The method of claim 1, wherein the image depicts concentration versus position of
- 35    one or more Golgi components within the cell.
8.     The method of claim 7, wherein the concentration of at least one Golgi component corresponds to intensity in the image.

9. The method of claim 1, wherein one or more cells in the image were treated with a material that binds to a component of the Golgi complex and emits a signal having an intensity corresponding to its concentration.

5

10. The method of claim 9 wherein the material is a lectin or antibody that binds to the component of the Golgi complex.

10

11. The method of claim 1, wherein the mathematical characterization of the Golgi complex comprises at least one of (i) an indicator of the peakedness of a histogram of at least one component of the Golgi complex, (ii) the texture of the Golgi complex and (iii) and the amount of Golgi complex in the region.

15

12. The method of claim 1, wherein the mathematical characterization of the Golgi complex comprises the kurtosis of intensity values obtained from the image.

20

13. The method of claim 1, wherein the mathematical characterization of the Golgi complex comprises a singular value decomposition of intensity values obtained from the image.

14. The method of claim 1, wherein the mathematical characterization of the Golgi complex comprises at least one of a mean and a standard deviation of intensity values obtained from the image.

25

15. The method of any one of claims 12, 13, and 14, wherein the intensity values correspond to local concentrations of a component of the Golgi complex within the single cell.

30

16. The method of claim 2, wherein classifying the Golgi complex comprises analyzing the mathematical characterization with a biological model of the Golgi complex.

17. The method of claim 16, wherein the biological model is a regression model or a neural network.

35

18. The method of claim 16, wherein the biological model is a classification and regression tree.

19. The method of claim 2, further comprising characterizing a population of cells from the image by considering the category of Golgi complex in each cell of the population.

5 20. The method of claim 19, further comprising predicting a mechanism of action from the characterization of the population.

21. A computer program product comprising a machine readable medium on which is provided program instructions for analyzing an image of one or more cells, the program  
10 instructions comprising:

(a) program code for identifying a region of the image subsuming some or all of the Golgi complex of a single cell;

(b) program code for automatically identifying, within said region, the location of the Golgi complex; and

15 (c) program code for using at least one of (i) the Golgi complex location within the region and (ii) the Golgi concentration within the region, to automatically mathematically characterize the Golgi complex within the single cell.

22. The computer program product of claim 21 further comprising:

20 (d) program code for using the mathematical characterization to automatically classify the Golgi complex in a category that at least partially distinguishes between normal Golgi and Golgi that is either diffuse or disperse or both diffuse and disperse.

23. The computer program product of claim 1, wherein the image comprises multiple  
25 cells, and wherein the program code for identifying the region of the image subsuming some or all of the Golgi complex comprises program code for segmenting the image into regions, each subsuming at least part of an individual cell.

24. The computer program product of claim 23, wherein the program code for  
30 segmenting comprises program code for identifying locations of nuclei in the cells.

25. The computer program product of claim 24, wherein the program code for identifying locations of nuclei comprises program code for identifying regions of the image where DNA is shown to concentrate.

35

26. The computer program product of claim 24, further comprising program code for dilating the locations of the nuclei within the image to subsume the locations of some or all of the Golgi complex in the individual cells.

5 27. The computer program product of claim 21, wherein the image depicts concentration versus position of one or more Golgi components within the cell.

10 28. The computer program product of claim 21, wherein the mathematical characterization of the Golgi complex comprises at least one of (i) an indicator of the peakedness of a histogram of at least one component of the Golgi complex, (ii) the texture of the Golgi complex and (iii) and the amount of Golgi complex in the region.

15 29. The computer program product of claim 21, wherein the mathematical characterization of the Golgi complex comprises at least one of (i) the kurtosis of intensity values obtained from the image, (b) a singular value decomposition of intensity values obtained from the image, (c) a mean and, (d) a standard deviation of intensity values obtained from the image.

20 30. The computer program product of claim 22, wherein the program code for classifying the Golgi complex comprises program code for analyzing the mathematical characterization with a biological model of the Golgi complex.

25 31. The computer program product of claim 30, wherein the biological model is a regression model or a neural network.

32. The computer program product of claim 30, wherein the biological model is a classification and regression tree.

30 33. The computer program product of claim 22, further comprising program code for characterizing a population of cells from the image by considering the category of Golgi complex in each cell of the population.

35 34. The computer program product of claim 33, further comprising program code for predicting a mechanism of action from the characterization of the population.

35. A method, implemented on a computing device, of producing a model for classifying cells based upon the condition of Golgi within the cells, the method comprising:

(a) receiving images of a plurality of cells of a training set, wherein individual cells of the training set have Golgi in various states;

(b) analyzing the images to mathematically characterize the Golgi within the multiple cells from the training set; and

(c) applying a modeling technique to the mathematical characterizations obtained in (b) to thereby produce the model.

36. The method of claim 35, further comprising segmenting at least one image of cells corresponding to some members of the training set to delineate individual cells within the image.

37. The method of claim 36, wherein the segmentation comprises dilating regions where the nuclei of the individual cells are found to reside.

38. The method of claim 35, wherein the images depict concentration versus position of one or more Golgi components within the cells.

39. The method of claim 38, wherein the concentration of at least one Golgi component corresponds to intensity in the images.

40. The method of claim 35, wherein the various states of Golgi in the cells of the training set are produced, at least in part, by treatment with multiple exogenous agents.

41. The method of claim 40, wherein the exogenous agents comprise at least one of drugs and drug candidates.

42. The method of claim 35, wherein the mathematical characterization of the Golgi comprises at least one of (i) an indicator of the peakedness of a histogram of at least one component of the Golgi, (ii) the texture of the Golgi and (iii) the amount of Golgi in the region.

43. The method of claim 35, wherein the mathematical characterization of the Golgi comprises one or more of the standard deviation, the mean, the kurtosis, and a singular value decomposition of intensity values obtained from individual cells of the image. The

method of claim 1, wherein the modeling technique includes at least one of a neural network, a regression technique, and a genetic algorithm.

5 44. The method of claim 35, wherein the modeling technique comprises generating a classification and regression tree.

45. A computer program product comprising a machine readable medium on which is provided program instructions for producing a model for classifying cells based upon the condition of Golgi within the cells, the program instructions comprising:

10 (a) program code for receiving images of a plurality of cells of a training set, wherein individual cells of the training set have Golgi in various states;

(b) program code analyzing the images to mathematically characterize the Golgi within the multiple cells from the training set; and

15 (c) program code for applying a modeling technique to the mathematical characterizations obtained in (b) to thereby produce the model

46. The computer program product of claim 45, further comprising program code for segmenting at least one image of cells corresponding to some members of the training set to delineate individual cells within the image.

20

47. The computer program product of claim 46, wherein the program code for segmentation comprises program code for dilating regions where the nuclei of the individual cells are found to reside.

25 48. The computer program product of claim 45, wherein the mathematical characterization of the Golgi comprises at least one of (i) an indicator of the peakedness of a histogram of at least one component of the Golgi, (ii) the texture of the Golgi and (iii) the amount of Golgi in the region.

30 49. The computer program product of claim 45, wherein the mathematical characterization of the Golgi comprises one or more of the standard deviation, the mean, the kurtosis, and a singular value decomposition of intensity values obtained from individual cells of the image. The method of claim 1, wherein the modeling technique includes at least one of a neural network, a regression technique, and a genetic algorithm.

35

50. The computer program product of claim 45, wherein the modeling technique comprises program code for generating a classification and regression tree.

51. An apparatus for automatically analyzing an image of one or more cells, the apparatus comprising:

an interface configured to receive the image of one or more cells;

5 a memory for storing, at least temporarily, some or all of the image; and

one or more processors in communication with the memory and designed or configured to segment the image into discrete regions, each subsuming some or all of the Golgi complex in single cell, and mathematically characterize the Golgi complex of single cells by operating on the discrete regions.

10 52. The apparatus of claim 51, wherein the one or more processors is further designed or configured to classify the Golgi complex based upon the mathematical characterization of the Golgi complex.

15 53. The apparatus of claim 52, wherein the one or more processors classifies the Golgi complex in a category that at least partially distinguishes between normal Golgi and Golgi that is diffuse or disperse.

20 54. The apparatus of claim 52, wherein the one or more processors classifies the Golgi complex by using a biological model.

55. The apparatus of claim 54, wherein the biological model is a classification and regression tree.

25 56. The apparatus of claim 51, wherein the one or more processors segment the image by first identifying regions of the image corresponding to nuclei of the one or more cells.

30 57. The apparatus of claim 56, wherein the one or more processors perform a dilation operation at the regions of the nuclei on the image in order to subsume the Golgi complex of each of the one or more cells.

58. The apparatus of claim 51, wherein the image to be received by the interface depicts concentration versus position of one or more Golgi components within the cell.

35 59. The apparatus of claim 58, wherein the one or more Golgi components correspond to intensity values in the image to be received by the interface.

60. The apparatus of claim 51, wherein the one or more cells in the image were treated with a material that binds to a component of the Golgi complex and emits a signal having an intensity corresponding to its concentration.

5 61. The apparatus of claim 51, wherein the one or more processors mathematically characterize the Golgi complex by calculating at least one or (i) an indicator of the peakedness of a histogram of at least one component of the Golgi complex, (ii) the texture of the Golgi complex, and (iii) the amount of Golgi complex in the discrete regions.

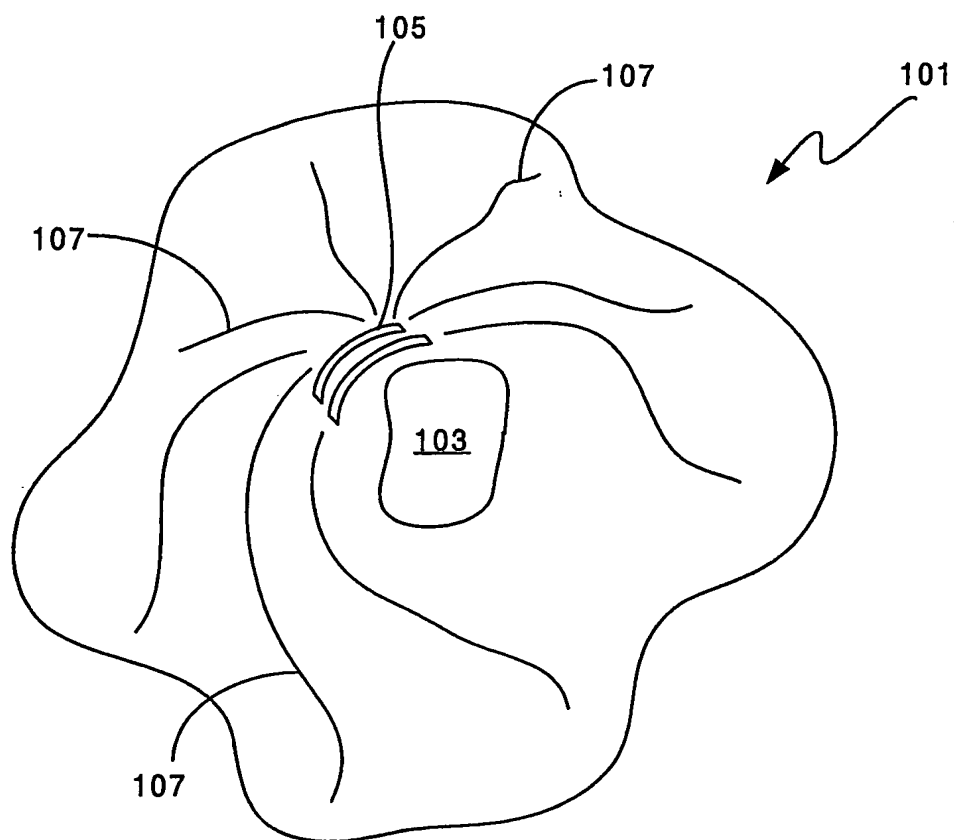
10 62. The apparatus of claim 51, wherein the one or more processors mathematically characterize the Golgi complex by calculating at least one of the kurtosis, the mean, the standard deviation, and a singular value decomposition of intensity values obtained from the discrete regions.

15 63. The apparatus of claim 51, wherein the one or more processors is further designed or configured to characterize a population comprising the one or more cells by considering a category of Golgi for each of the one or more cells.

20 64. The apparatus of claim 63, wherein the one or more processors is further designed or configured to predict a mechanism of action from characterizing the population.

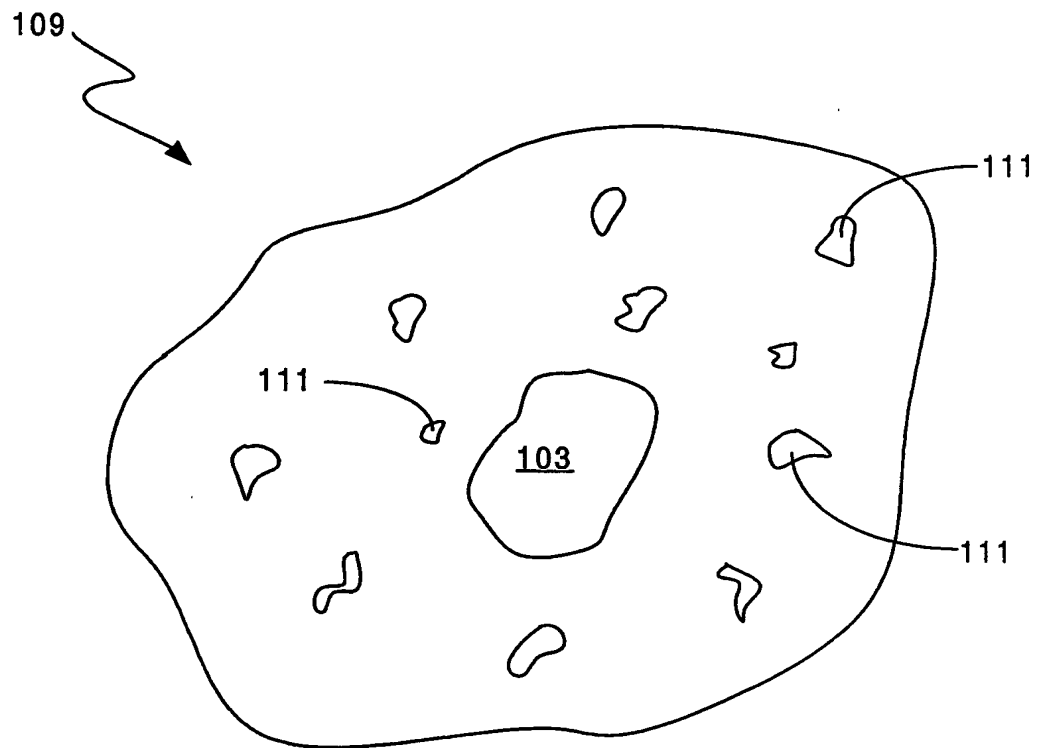


1/13



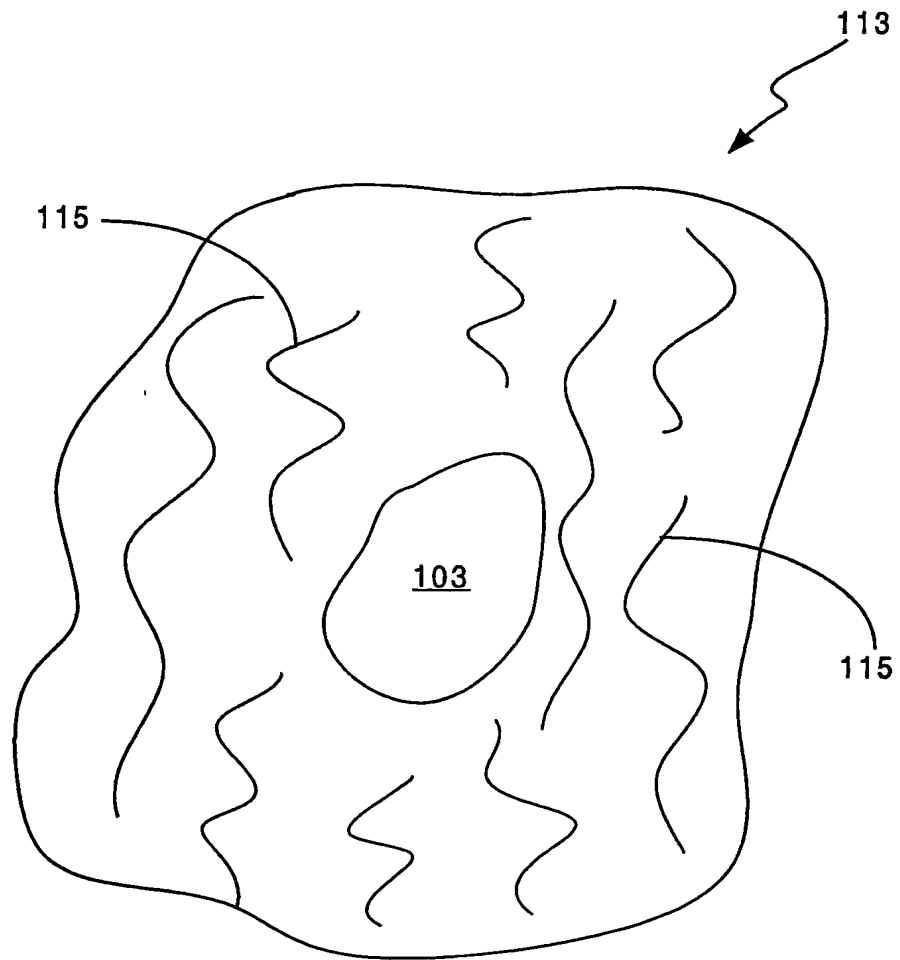
**FIG. 1A**

2/13



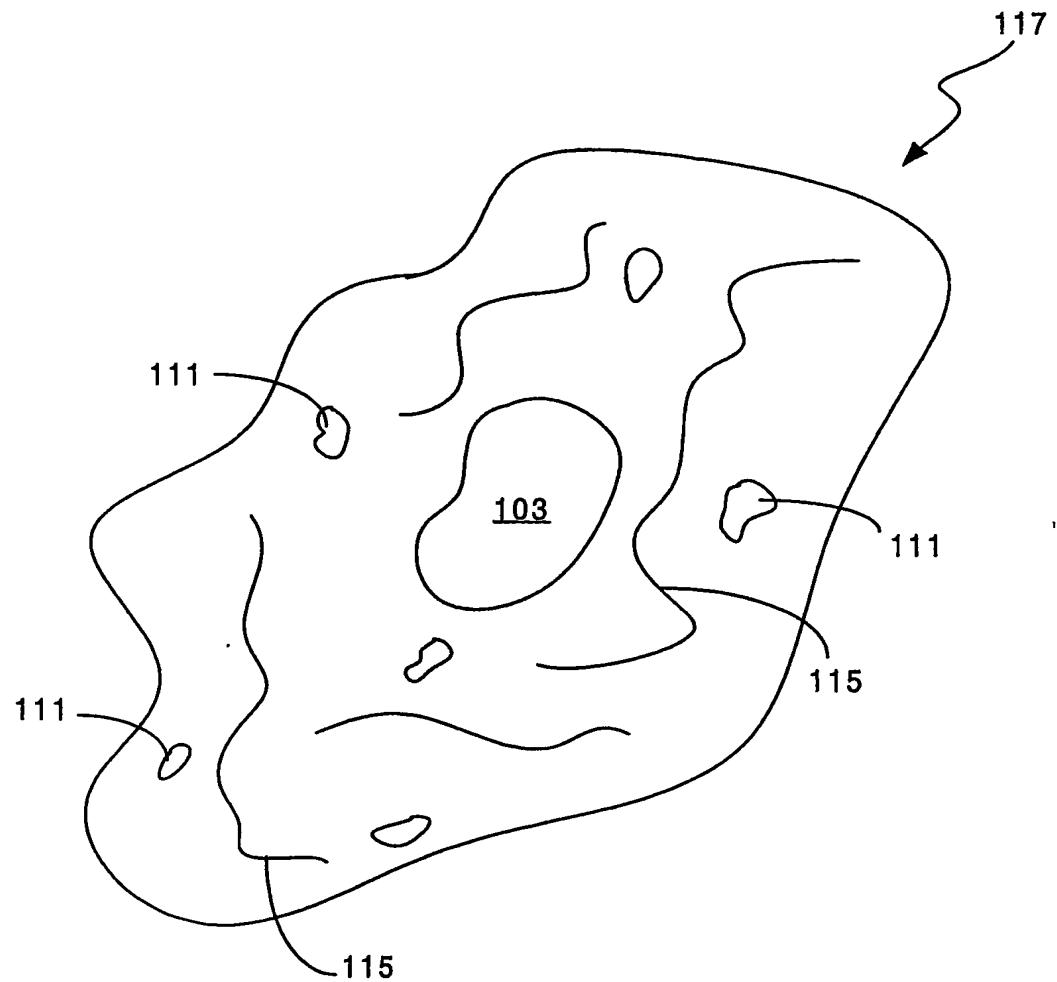
**FIG. 1B**

3/13



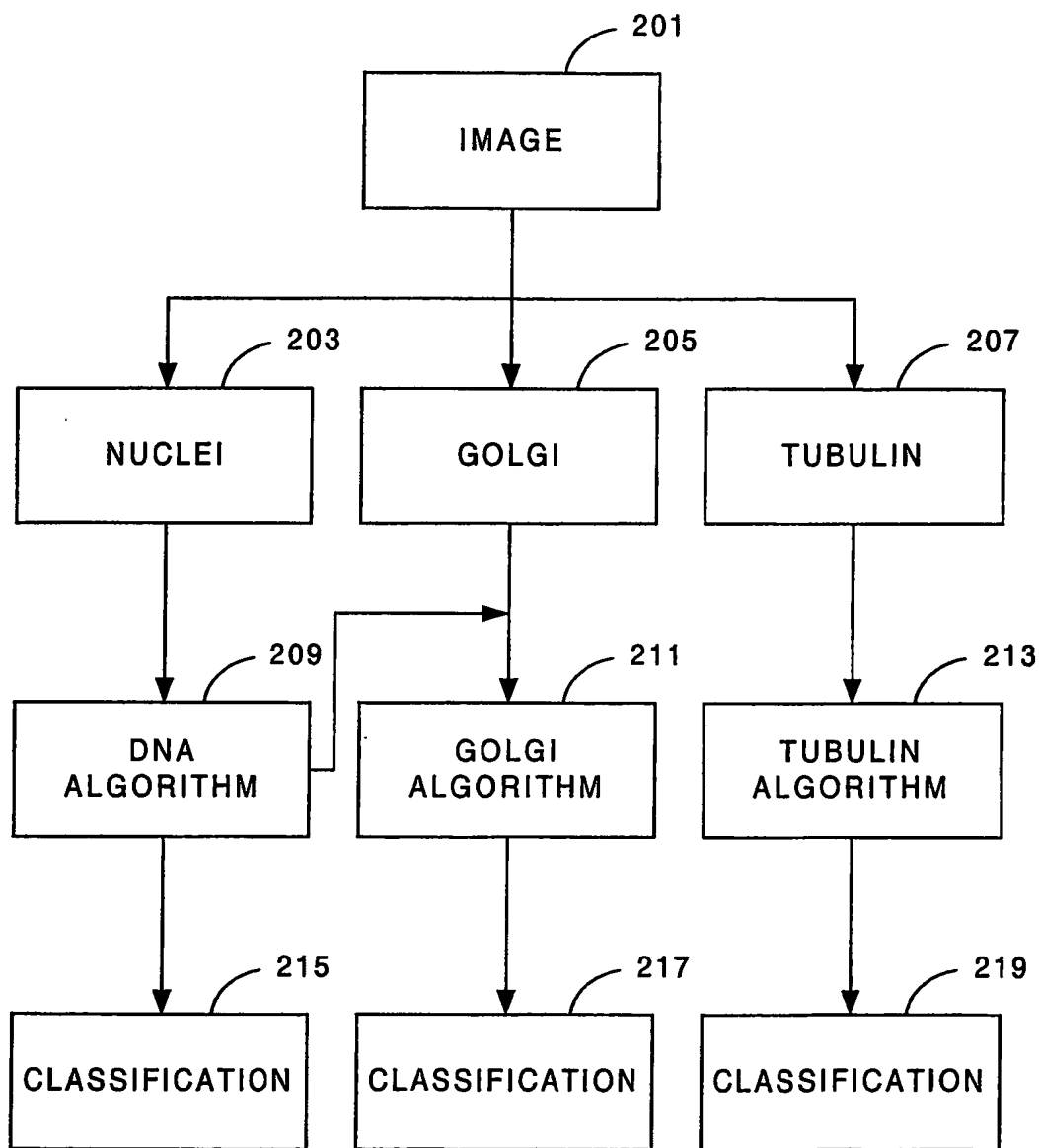
**FIG. 1C**

4/13

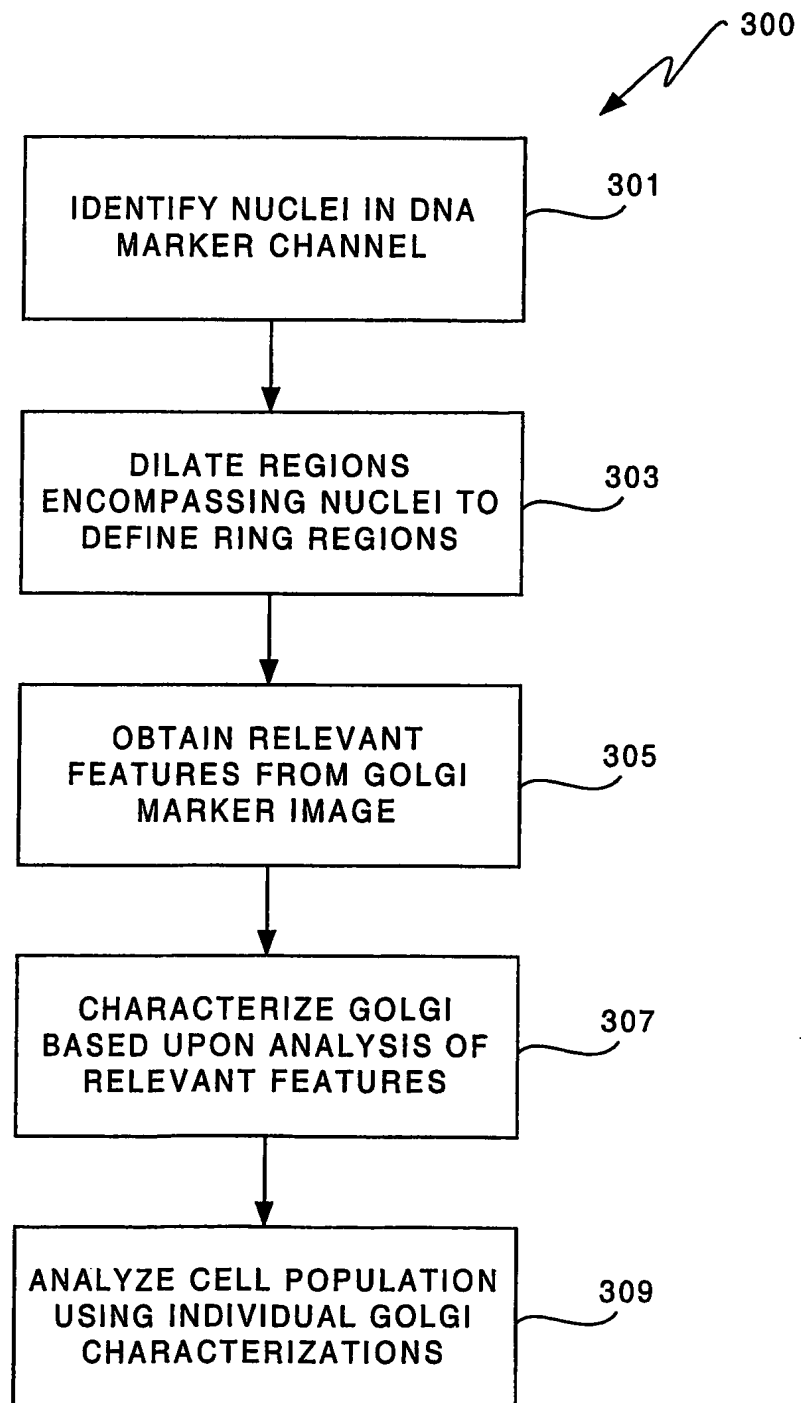


**FIG. 1D**

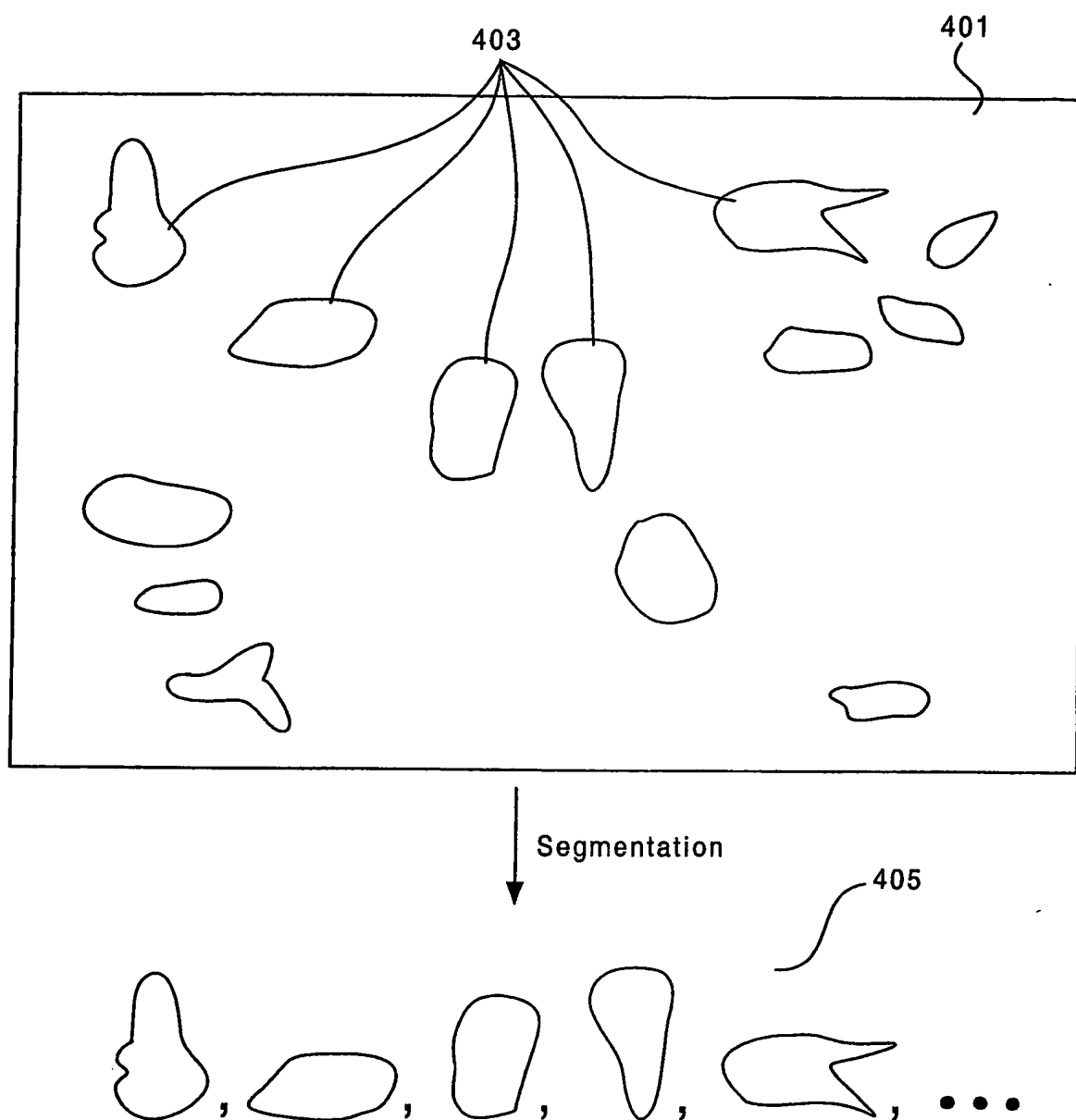
5/13

**FIG. 2**

6/13

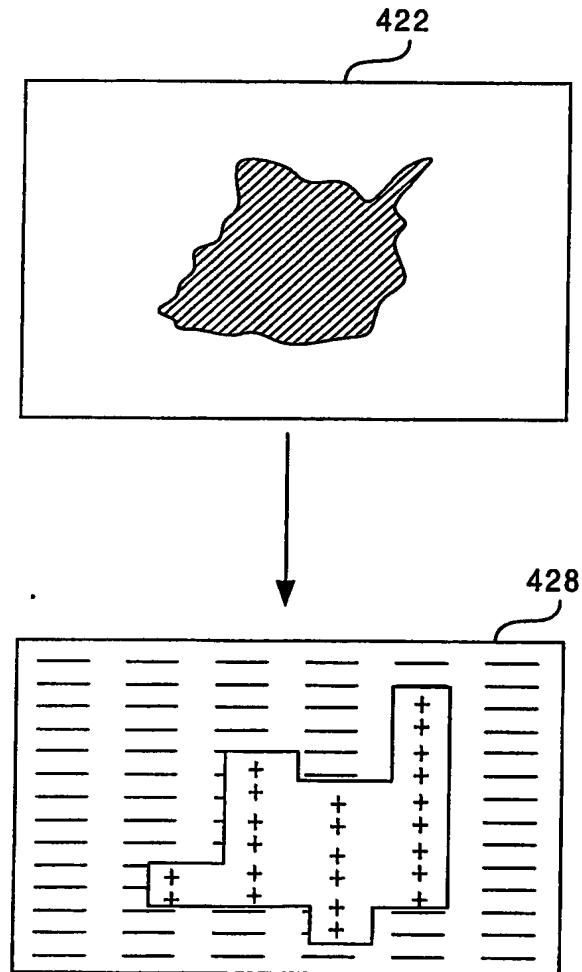
**FIG. 3**

7/13



**FIG. 4A**

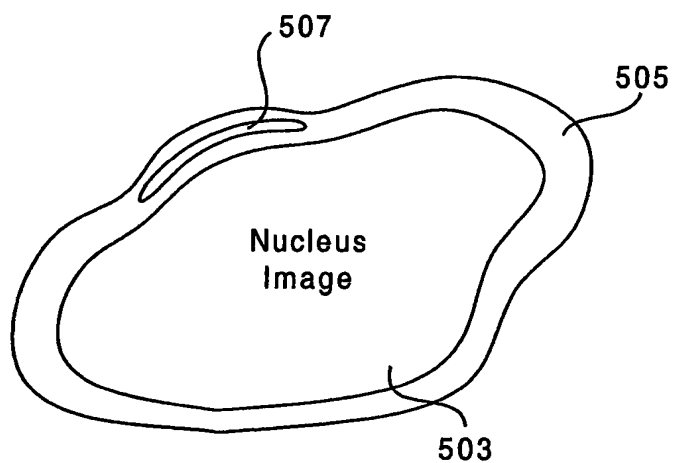
8/13



**FIG. 4B**



9/13

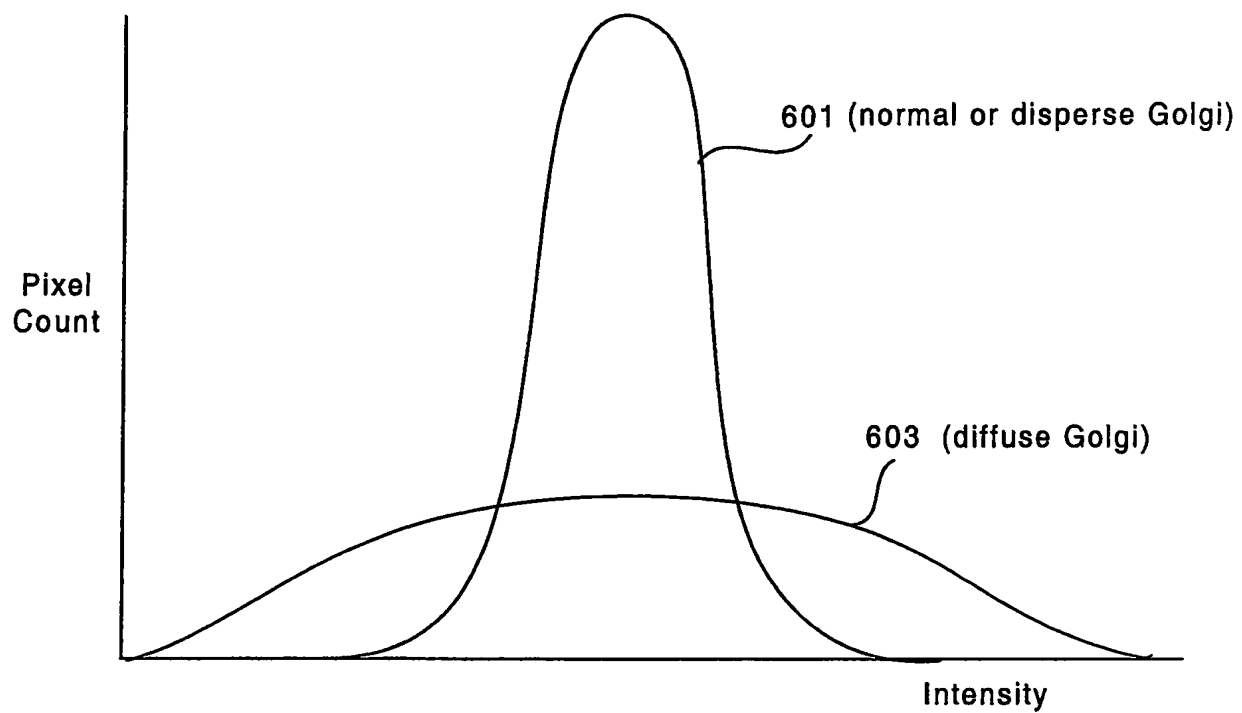
**FIG. 5A**

						O	O
					O	X	X
				O	X	N	N
			O	X	N	N	N
		O	X	N	N	N	N
	O	X	N	N	N	N	N

511

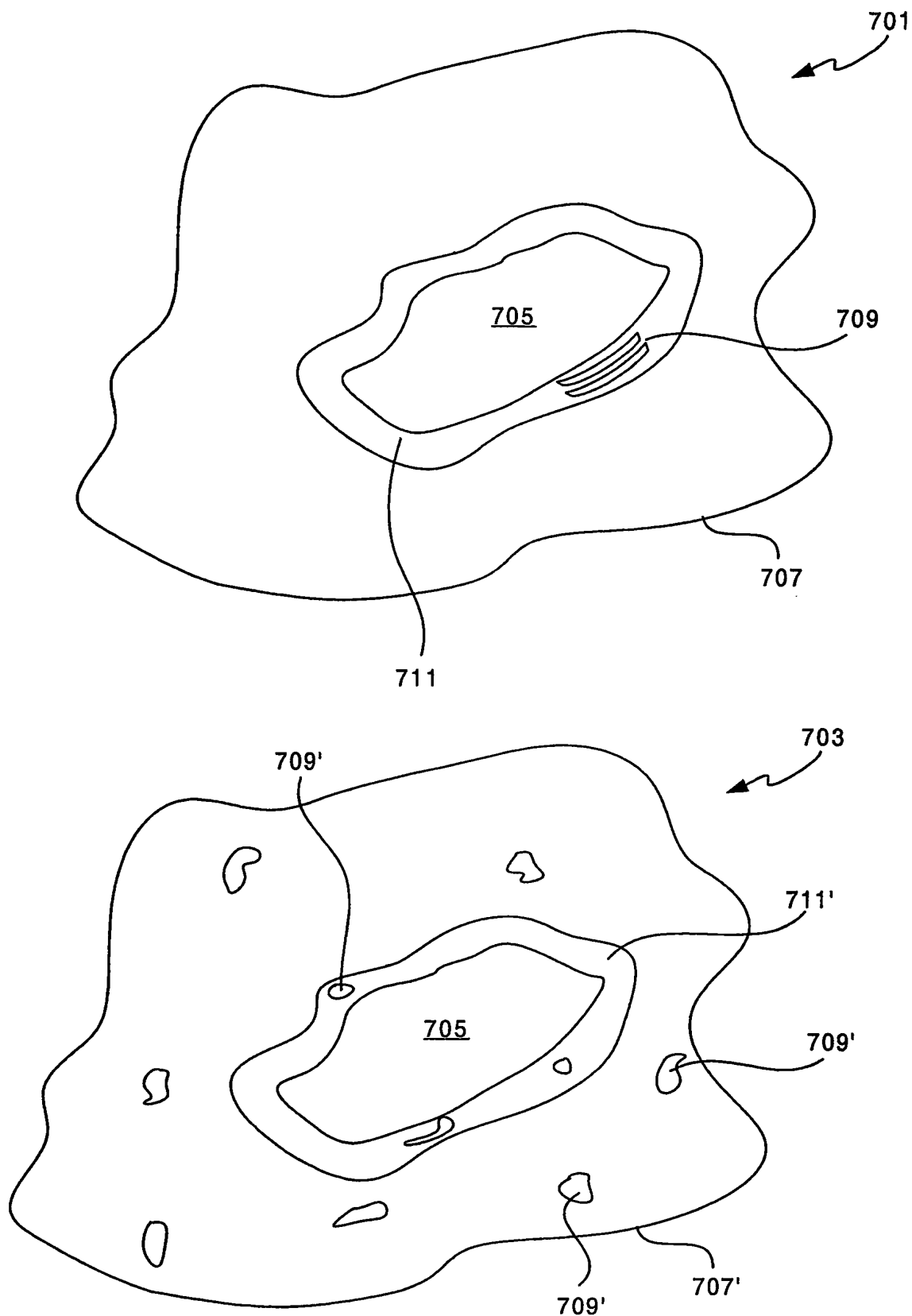
**FIG. 5B**

10/13



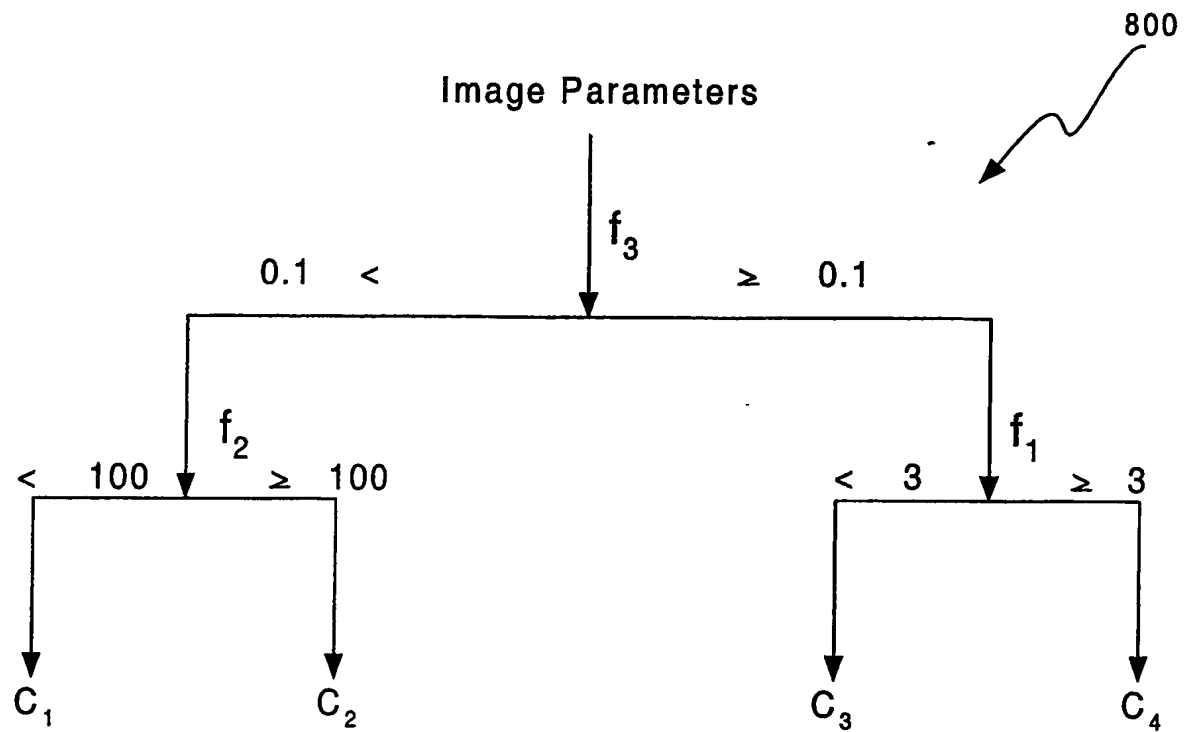
**FIG. 6**

11/13

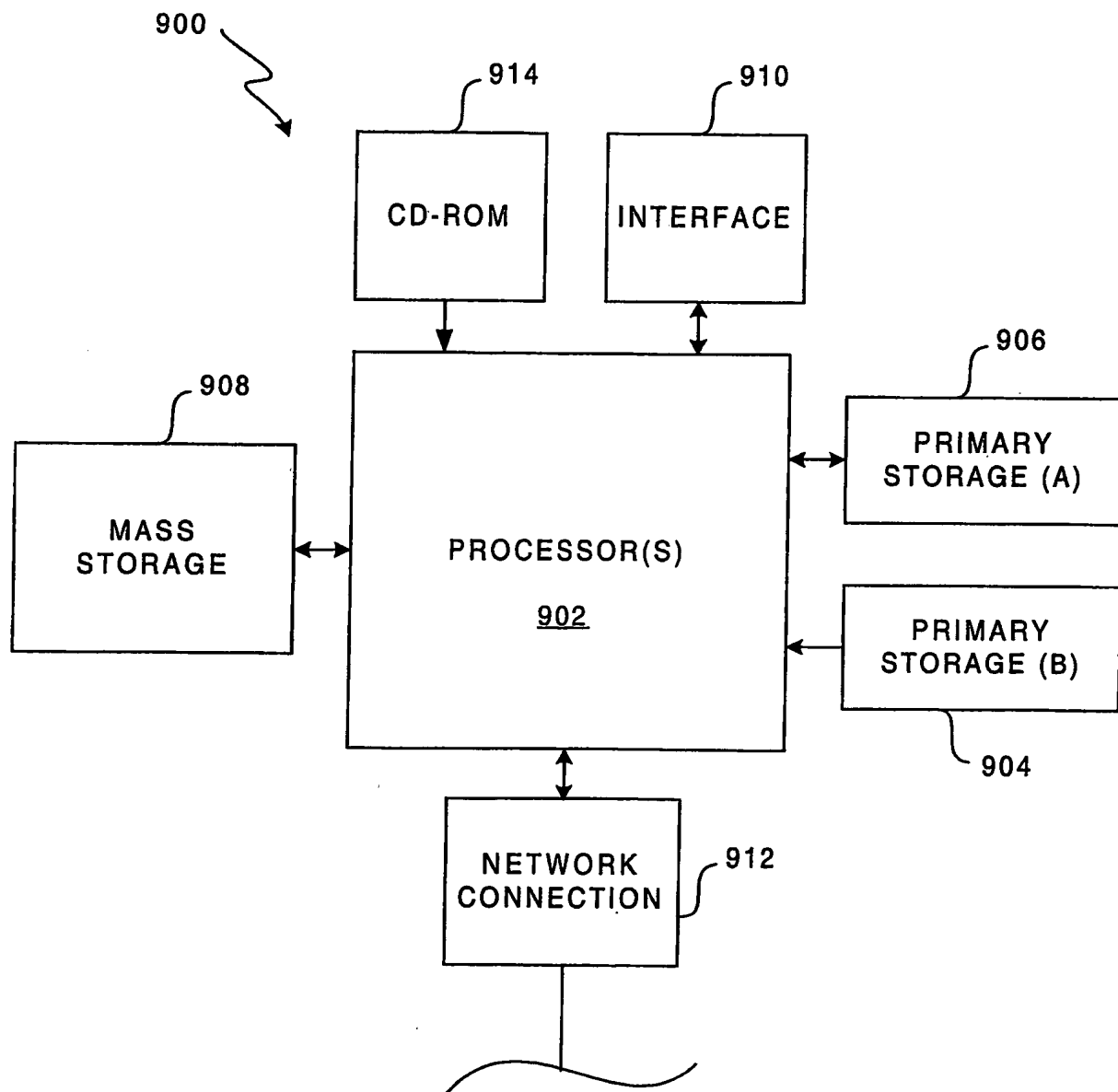


**FIG. 7**

12/13

**FIG. 8**

13/13

**FIG. 9**

# INTERNATIONAL SEARCH REPORT

inal Application No  
PCT/US 02/04492

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 G01N33/50 G01N15/14 G06K9/00 G06K9/46

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G01N G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 00 70528 A (CYTOKINETICS INC) 23 November 2000 (2000-11-23) page 9, line 3-5 page 65, line 31,32	1-64
Y	US 5 548 661 A (GOUGH DAVID ET AL) 20 August 1996 (1996-08-20) column 19, line 32-54 ----- -/-	1-64

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

### \* Special categories of cited documents:

\*A\* document defining the general state of the art which is not considered to be of particular relevance

\*E\* earlier document but published on or after the International filing date

\*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

\*O\* document referring to an oral disclosure, use, exhibition or other means

\*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\*G\* document member of the same patent family

Date of the actual completion of the international search

26 September 2002

Date of mailing of the international search report

08/10/2002

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax (+31-70) 340-3016

Authorized officer

Müller, T

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 02/04492

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>SERBOUTI S ET AL: "Image Segmentation And Classification Methods To Detect Leukemias" ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY, 1991. VOL.13: 1991., PROCEEDINGS OF THE ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ORLANDO, FL, USA 31 OCT.-3 NOV. 1991, NEW YORK, NY, USA, IEEE, US, 31 October 1991 (1991-10-31), pages 260-261, XP010101605 ISBN: 0-7803-0216-8 the whole document</p>	1-64
Y	<p>CSEKE I: "A fast segmentation scheme for white blood cell images" PATTERN RECOGNITION, 1992. VOL.III. CONFERENCE C: IMAGE, SPEECH AND SIGNAL ANALYSIS, PROCEEDINGS., 11TH IAPR INTERNATIONAL CONFERENCE ON THE HAGUE, NETHERLANDS 30 AUG.-3 SEPT. 1992, LOS ALAMITOS, CA, USA, IEEE COMPUT. SOC, US, 30 August 1992 (1992-08-30), pages 530-533, XP010030223 ISBN: 0-8186-2920-7 the whole document</p>	1-64

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 02/04492

Patent document cited in search report		Publication date		Patent family member(s)		Publication date
WO 0070528	A	23-11-2000	AU	4847100 A		05-12-2000
			EP	1188139 A2		20-03-2002
			WO	0070528 A2		23-11-2000
US 5548661	A	20-08-1996	US	5790692 A		04-08-1998
			US	5856665 A		05-01-1999
			US	5790710 A		04-08-1998



(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
29 August 2002 (29.08.2002)

PCT

(10) International Publication Number  
**WO 02/067188 A3**

(51) International Patent Classification<sup>7</sup>: **G01N 33/50**,  
15/14, G06K 9/00, 9/46

(21) International Application Number: PCT/US02/04492

(22) International Filing Date: 14 February 2002 (14.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
09/792,012 20 February 2001 (20.02.2001) US

(71) Applicant (*for all designated States except US*): **CY-TOKINETICS, INC.** [US/US]; 280 E. Grand Avenue, Suite 2, South San Francisco, CA 94080 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **CONG, Ge** [CN/US]; 10944 San Pablo Avenue, Apartment 226, El Cerrito, CA 94530 (US). **VAISBERG, Eugeni, A.** [US/US]; 647 Pegasus Lane, Foster City, CA 94404 (US). **WU, Hsien-Hsun** [CN/US]; 1764 Wayne Circle, San Jose, CA 95131 (US).

(74) Agent: **WEAVER, Jeffrey, K.**; Beyer Weaver & Thomas, LLP, P.O. Box 778, Berkeley, CA 94704 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

(88) Date of publication of the international search report:  
10 April 2003

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: IMAGE ANALYSIS OF THE GOLGI COMPLEX

(57) Abstract: Methods, code and apparatus analyze cell images to automatically identify and characterize the Golgi complex in individual cells. This is accomplished by first locating the cells in the image and defining boundaries of those cells that subsume some or all of the Golgi complex of those cells. The Golgi complex in the images typically have intensity values corresponding to the concentration of a Golgi component in the cell (e.g. a polysaccharide associated with the Golgi complex). The method/system then analyzes the Golgi components of the image (typically on a pixel-by-pixel basis) to mathematically characterize the Golgi complex of individual cells. This mathematical characterization represents phenotypic information about the cells' Golgi complex and can be used to classify cells. From this information, mechanism of action and other important biological information can be deduced.



WO 02/067188 A3

## INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 02/04492

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 G01N33/50 G01N15/14 G06K9/00 G06K9/46

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G01N G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the International search (name of data base and, where practical, search terms used)

EPO-Internal

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 00 70528 A (CYTOKINETICS INC) 23 November 2000 (2000-11-23) page 9, line 3-5 page 65, line 31,32	1-64
Y	US 5 548 661 A (GOUGH DAVID ET AL) 20 August 1996 (1996-08-20) column 19, line 32-54 -/-	1-64

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

## \* Special categories of cited documents:

\*A\* document defining the general state of the art which is not considered to be of particular relevance

\*E\* earlier document but published on or after the international filing date

\*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

\*O\* document referring to an oral disclosure, use, exhibition or other means

\*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\*&amp;\* document member of the same patent family

Date of the actual completion of the international search

26 September 2002

Date of mailing of the international search report

08/10/2002

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Müller, T

# INTERNATIONAL SEARCH REPORT

II International Application No  
PCT/US 02/04492

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>SERBOUTI S ET AL: "Image Segmentation And Classification Methods To Detect Leukemias"</p> <p>ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY, 1991. VOL.13: 1991., PROCEEDINGS OF THE ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ORLANDO, FL, USA 31 OCT.-3 NOV. 1991, NEW YORK, NY, USA, IEEE, US, 31 October 1991 (1991-10-31), pages 260-261, XP010101605</p> <p>ISBN: 0-7803-0216-8</p> <p>the whole document</p> <p>---</p>	1-64
Y	<p>CSEKE I: "A fast segmentation scheme for white blood cell images"</p> <p>PATTERN RECOGNITION, 1992. VOL.III. CONFERENCE C: IMAGE, SPEECH AND SIGNAL ANALYSIS, PROCEEDINGS., 11TH IAPR INTERNATIONAL CONFERENCE ON THE HAGUE, NETHERLANDS 30 AUG.-3 SEPT. 1992, LOS ALAMITOS, CA, USA, IEEE COMPUT. SOC, US, 30 August 1992 (1992-08-30), pages 530-533, XP010030223</p> <p>ISBN: 0-8186-2920-7</p> <p>the whole document</p> <p>-----</p>	1-64

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 02/04492

Patent document cited in search report		Publication date		Patent family member(s)	Publication date
WO 0070528	A	23-11-2000	AU	4847100 A	05-12-2000
			EP	1188139 A2	20-03-2002
			WO	0070528 A2	23-11-2000
US 5548661	A	20-08-1996	US	5790692 A	04-08-1998
			US	5856665 A	05-01-1999
			US	5790710 A	04-08-1998